ESnet4: Support for TEAM Portals

Eli Dart
Network Engineer
Network Engineering Group

Energy Sciences Network
Lawrence Berkeley National Laboratory

June 23, 2008

Networking for the Future of Science

Overview

- **ESnet defined**

- **Scope of Collaboration and Connectivity**

- **ESnet architecture and network services**

- **ESnet connections to site networks**

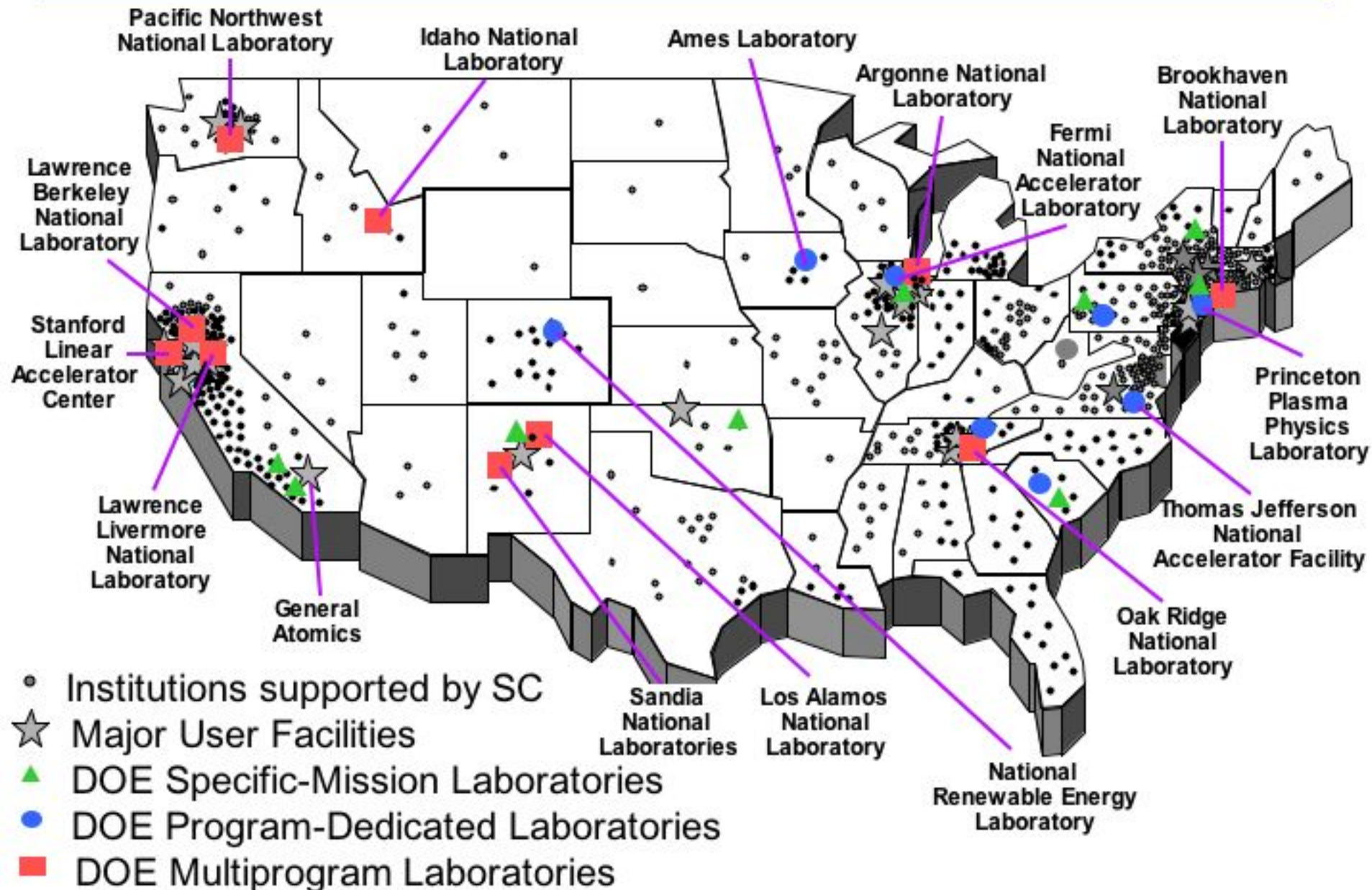- **Networking for TEAM portals**

- **Questions**

ESnet Defined

- **ESnet is the high performance networking facility of the DOE Office of Science**

- **ESnet's primary mission is to enable the large-scale science that is the mission of the DOE Office of Science**
  - Structure of modern scientific endeavor is highly collaborative and distributed – many sites, many institutions, many countries
  - This structure assumes the existence of a high performance, feature-rich network fabric that interconnects all components of a collaboration
  - Massive data sets, remote data management, visualization, etc.

- **ESnet is a nationwide network that provides connectivity to the DOE National Laboratories**
  - High-bandwidth peerings with other national and international networks such as I2, GEANT, CANARIE, NLR, Kreonet2, SINET
  - Full connectivity to the commodity Internet in many locations

Scope of Collaboration and Connectivity

- **DOE Office of Science collaborations typically span National Laboratories and US/International universities**

- **Bandwidth, Connectivity and Services must span multiple high-performance networks**

  – ESnet routinely collaborates with other major R&E networks such as Internet2, GEANT, and CANARIE to ensure interoperability of advanced services

  – Scope of collaboration drives scope of required connectivity

- **Question for TEAM collaboration – where are your users?**

  – Plans for portals are at DOE facilities

  – What about taking the data home?

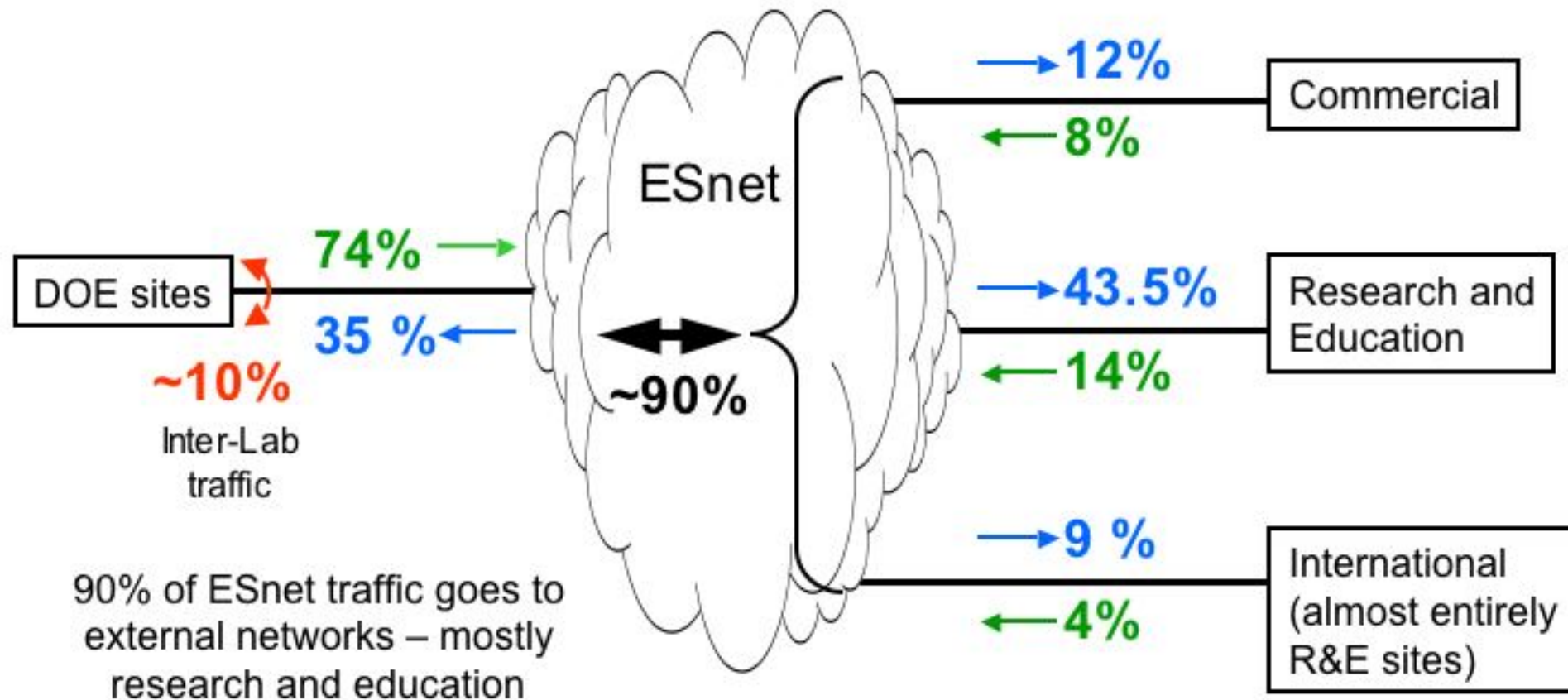 – 800MB/sec can amount to a lot of data very quickly

The US Collaborators of DOE's Office of Science, Drives ESnet Design for Domestic Connectivity

Pacific Northwest National Laboratory

Idaho National Laboratory

Ames Laboratory

Argonne National Laboratory

Fermi National Accelerator Laboratory

Brookhaven National Laboratory

Lawrence Berkeley National Laboratory

Stanford Linear Accelerator Center

Lawrence Livermore National Laboratory

General Atomics

Princeton Plasma Physics Laboratory

Thomas Jefferson National Accelerator Facility

Oak Ridge National Laboratory

Sandia National Laboratories

Los Alamos National Laboratory

National Renewable Energy Laboratory

- ○ Institutions supported by SC
- ☆ Major User Facilities
- ▲ DOE Specific-Mission Laboratories
- ● DOE Program-Dedicated Laboratories
- ■ DOE Multiprogram Laboratories

Where Does ESnet Traffic Originate and Terminate?

ESnet Inter-Sector Traffic Summary, Jan. 2007

Traffic coming into ESnet = Green, Traffic leaving ESnet = Blue, Traffic between ESnet sites = Red

ESnet

→ **12%** Commercial
← **8%**

DOE sites
74% →
35 % ←

~10%
Inter-Lab traffic

~90%

→ **43.5%** Research and Education
← **14%**

90% of ESnet traffic goes to external networks – mostly research and education

→ **9 %** International (almost entirely R&E sites)
← **4%**

➢ **Clearly the design and operation of the peering points where ESnet passes traffic to other research and education networks is as important as the design and operation of the core network**
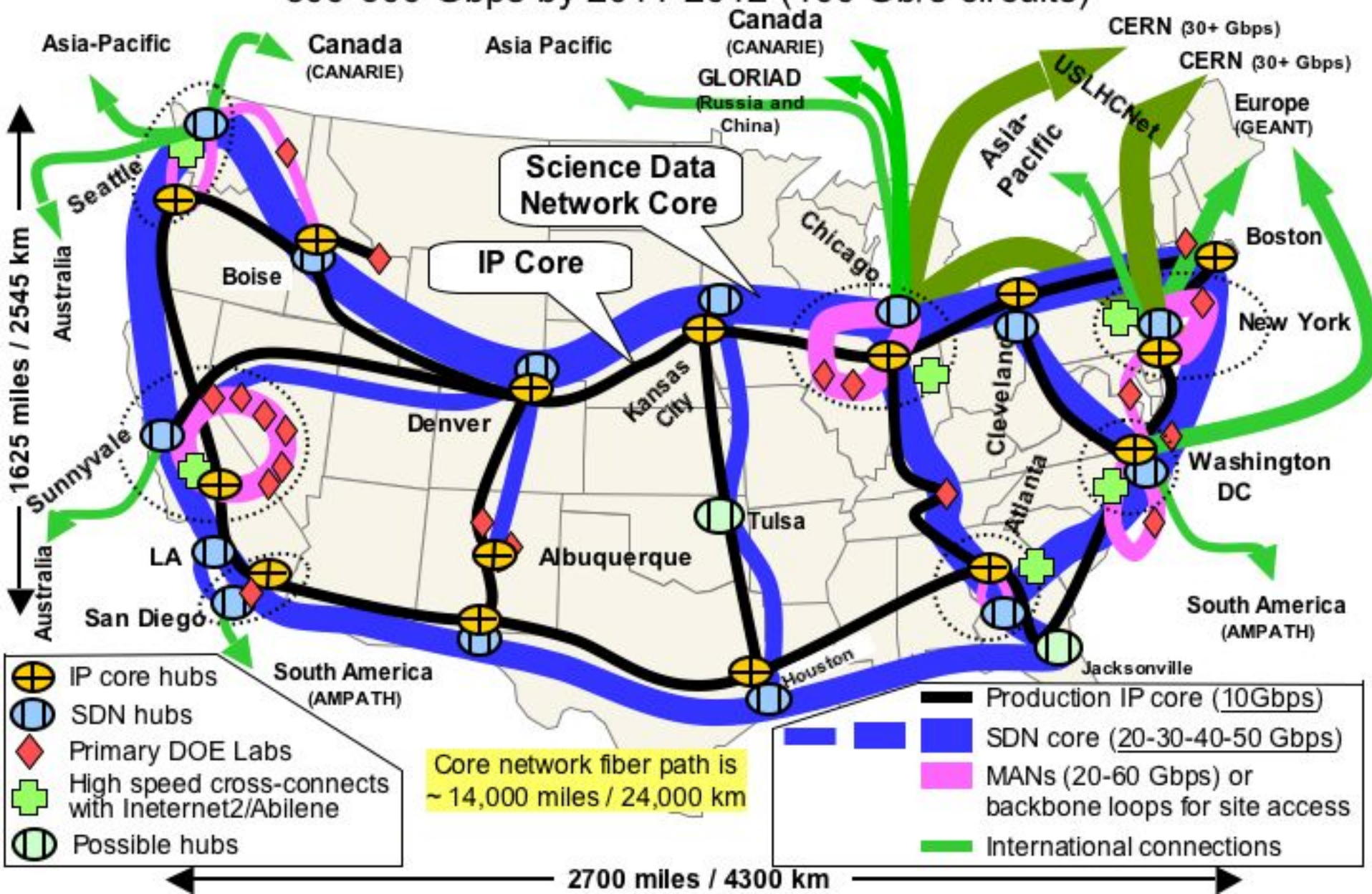
ESnet Architecture – ESnet4

- **ESnet is built on a nationwide circuit infrastructure**

- **ESnet4 – new architecture to serve both general-purpose networking needs and large-scale science**
  - Two networks – IP and SDN
    - IP network serves general-purpose networking needs for laboratory business, low-bandwidth (up to 1Gbps or so) scientific applications, connectivity to Google, Yahoo, etc.
    - SDN – Science Data Network is built specifically to serve the needs of large-scale science → guaranteed bandwidth, traffic isolation, bandwidth reservation
  - IP network will remain 10Gbps for the foreseeable future
  - SDN will see significant additional capacity over the next few years

ESnet Provides Global High-Speed Internet Connectivity for DOE Facilities and Collaborators (12/2007)
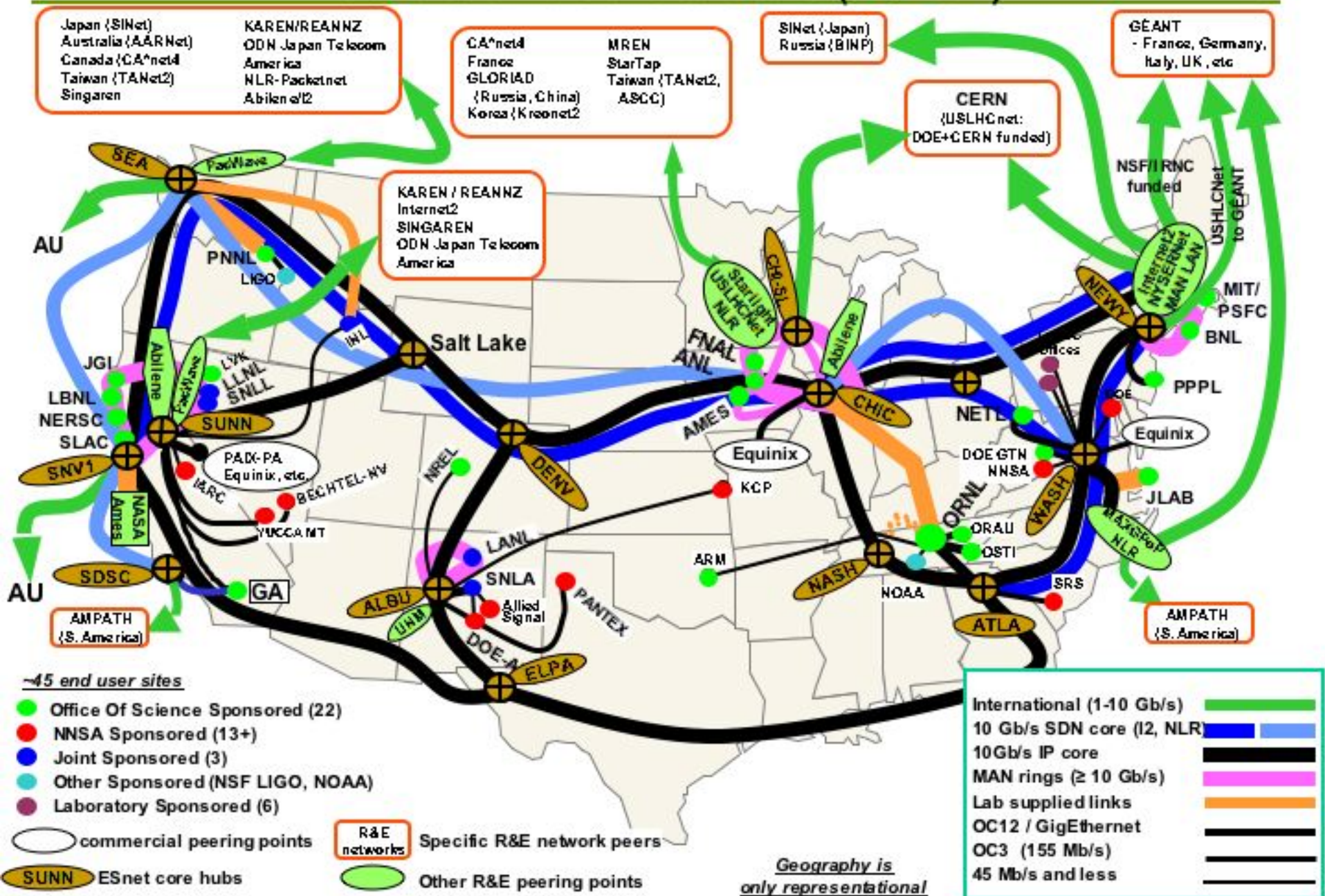
ESnet4 End-Game

Core networks 50-60 Gbps by 2009-2010 (10Gb/s circuits), 500-600 Gbps by 2011-2012 (100 Gb/s circuits)

Legend:
- **IP core hubs**
- **SDN hubs**
- **Primary DOE Labs**
- **High speed cross-connects with Ineternet2/Abilene**
- **Possible hubs**

Core network fiber path is ~ 14,000 miles / 24,000 km

- Production IP core (10Gbps)
- SDN core (20-30-40-50 Gbps)
- MANs (20-60 Gbps) or backbone loops for site access
- International connections

1625 miles / 2545 km

2700 miles / 4300 km

Network Services

- **In addition to the traditional best-effort routed service of the IP network, ESnet4 provides additional services through SDN:**

  – Virtual Circuit service through OSCARS (http://www.es.net/OSCARS/index.html)

  – Virtual Circuits provide bandwidth guarantees, traffic isolation

  – Scheduled bandwidth

  – Typical demarcation point for a Virtual Circuit is a VLAN tag at the site network connection to ESnet

- **SDN virtual circuits are currently being used in production by the High Energy Physics community for moving LHC data**

- **Site connection to SDN is typically via a second router or switch interface**

Site Connections to ESnet

- **National Laboratory site networks are the gateways to ESnet services**

  - ESnet works in partnership with site networks to deliver services to scientists

  - Networking is inherently end-to-end – all parts must work together to provide service

    - End hosts

    - Site Networks

    - Core Networks

- **Site networks must be involved in planning for new capabilities**

  - We can't do it alone – the sites are an integral part of our success and the success of our scientific constituents

  - Early discussions and planning avoid delays (e.g. if new fiber optic cables must be installed between buildings to enable higher-bandwidth connectivity)

Networking for TEAM Portals

- **TEAM has one instrument that will require high bandwidth end to end connectivity**

  – Main instrument data rate is 800MB/sec → 6.4Gbps (bulk data for TCP transfer…disk to disk)

  – FluCam is 30MB/sec → 240Mbps – not so bad

  – At the moment, this should fit in a 1Gig pipe

- **TEAM portal may be a good candidate for SDN virtual circuits**

  – Bandwidth guarantees

  – Traffic isolation

- **Site networking issues to be considered**

  – Bandwidth guarantees in site networks

  – Firewalls don't always perform well (e.g. per-flow bandwidth limitations)

  – Work together with site networking personnel

Network Tuning for Bulk Data Movement

- **There are several common problems associated with bulk data movement**

- **Poor performance is often the result of un-tuned hosts or inefficient protocols**

  - Tune your machines – defaults rarely work well for high performance bulk data transfer
    - TCP tuning
    - Make sure the application uses tuned parameters
  - Use bulk data transfer tools that are optimized for the WAN
  - GridFTP and bbcp both use TCP tuning and parallel streams
  - DO NOT USE scp/sftp – these use libssl library with internal buffering method that does not work well on "long fat" networks

- **DOE sites should be getting good performance**
  - 1TB in 8 hours is approximately 300Mbps
  - If you can't achieve this, and disk is not the bottleneck, something's broken

- **More info at http://fasterdata.es.net**

Questions?

- **Thanks!**